

내용기반 영상 검색 유효성을 측정하는 방법들에 대한 비교 (Precision/Recall vs. Wilcoxon 순위 방법)

장 순자 김 형중 여 인권
*강원대학교 전기·전자·정보통신공학부

Comparison of Retrieval Effectiveness between Precision/Recall and Wilcoxon Test

Jang, Soon-Ja Kim, Hyoung-Joung Yeo, In-Kwon
Division of Electrical and Electronics Engineering, Kangwon University
E-mail: jsj@multimedia.kangwon.ac.kr, khj@cc.kangwon.ac.kr, inkwon@dragon.skku.ac.kr

요 약

영상검색은 문자위주의 검색의 한계로, 그 내용에 기반 한 특징벡터를 이용하여 검색을 수행한다. 특징벡터간의 거리를 계산하고 그 값들에 순위를 매긴다. 이렇게 얻어진 순위 값들을 가지고 사용된 검색방법의 유효성을 검사하는데 Recall/Precision방법이 이용되고 있다. 질의영상과 같은 군에 속하는 영상이 얼마나 검출되었는지를 검출된 영상에 기반 하여 계산하거나, 영상 군에 기반 하여 계산하는 방법들이다. 그러나, 검출되는 순위 값의 범위를 정하고, 그 범위 내에 속하는 만족하는 값의 개수를 세는 방법을 이용한다. 따라서 주어진 두 샘플의 전체적인 경향을 비교하지는 못한다. 본 연구에서는 순위를 이용하여 비교하고자하는 두 샘플의 순위들을 결합하여 순위를 매기고 각 샘플들에 매겨진 순위값들의 평균과 분산을 이용하여 각 샘플들을 전체적으로 비교할 수 있다.

1. 서론

영상 검색은 영상 데이터베이스 내에 있는 영상들을 비슷한 것끼리 묶고서, 질의 영상이 주어졌을 때 데이터베이스 내에서 질의 영상과 유사한 영상 군을 검색하는 방법이다. 내용기반 영상 검색은 영상의 내용을 기반으로 얻어진 특징벡터를 비교하는 방법으로, 특징벡터간의 거리를 이용하여 비교할 영상간의 유사도를 측정한다. 여기서, 사용된 특징벡터 추출 방법들이 얼마나 유효한지를 측정할 필요가 있다.

대부분의 논문들에서는, 특징벡터 추출 방법의 효율성을 비교하기 위해 Recall/Precision 방법을 이용하고 있다 [1][3][4][5][6][7][8]. 이 방법은 특징벡터간의 거리 값들을 올림차순으로 정렬한 후에 특정 개수 내에서 검출된 질의 영상과 관계 있는 영상의 비율(Precision)과 질의 영상이 속하는 영상군중 특정 개수 내에서 검출된 영상의 비율(Recall)을 비교하여 검색방법의 효율성을 판정한다. 그러나 이 방법은 특정 개수에 따라 결과가 달라질 수 있으며 정확한 결과 분석을 위한 특정 개수를 정하는 것이 어렵

다. 이러한 문제점을 해결하기 위해 [1]과 [8]에서는 개선된 Recall/Precision 방법을 사용하였다. [8]에서는 Methref[9]의 방법인

$$\text{effectiveness} = \text{Recall} \quad (\text{특정개수} > N) \\ \text{Precision} \quad (\text{특정개수} \leq N)$$

을 이용한다. 여기서 N은 데이터베이스 내에서 질의영상과 관계 있는 영상들의 총 개수이다. 이 방법 역시 특정개수의 영향을 받으며, 전체적인 효과를 평가하지 못한다. [1]에서는 주어진 질의 영상 내, object들간의 거리 임계치를 주고, ideal system에 의해 얻어진 model object들이 데이터베이스 내에서 모두 관계 있는 것들이라고 가정된 후, 상대적인 Recall(rPr)과 상대적인 Precision(rPc)을 구한다.

$$rPr = \frac{\text{number of relevant objects retrieved by test system}}{\text{total number of objects retrieved}} \\ rPc = \frac{\text{number of relevant objects retrieved by test system}}{\text{total number of objects retrieved by ideal system}}$$

여기서, 임계 치가 클 때 향상된 결과를 보이지만, 역시 임계 치 범위 내에 존재하는 영상의 개수인, 특정개수에 영향을 받는다.

본 논문에서는 검색 순위를 이용하여 검색 방법의 효율성을 비교하는 방법을 제안한다. 그리고, 이 방법과 Recall/Precision 방법을 비교 고찰한다.

2. Recall / Precision

영상 검색에 사용된 방법들의 효율성을 정량적으로 비교 평가하기 위해 주로 사용되는 방법이 Recall/Precision 값이다. 이 방법을 사용하기 위해서는 일단 데이터 베이스 내에 있는 영상들을 검색에 앞서 미리 비슷한 것끼리 분류해 놓아야한다. 그런 다음, 질의 영상이 주어지면, 한 특징벡터 추출 방법을 이용하여 데이터베이스 내에 있는 모든 영상의 특징벡터를 계산하고, 특징벡터들간의 거리들을 구한 다음 올림차순으로 이 값들을 정렬한다. 이렇게 얻어진 값들 중에서, Recall/Precision의 비율을 계산하기 위해 고려할 대상의 개수를 결정해야 하는데 이 때 고려된 대상의 개수를 특정 개수(scope)라고 한다. 그림 1에서 A는 올림차순으로 표시된 자료들 중 특정 개수 범위 내에 들어 있는 영상들을 나타내고 B는 질의 영상과 비슷하다고 미리 분류된 영상들을 나타내며 C(A ∩ B)는 영상검색의 결과로 특정 개수 범위에 속하는 영상이면서 질의 영상과 관계가 있다고 미리 분류된 영상들을 나타낸다.

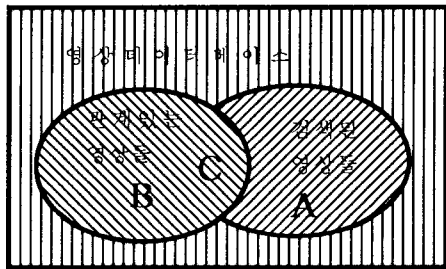


그림 1. Recall과 Precision

여기서 Recall과 Precision을 수식으로 표현하면 다음과 같다[11].

$$\text{Recall} = P(A|B) = P(A \cap B) / P(B)$$

$$\text{Precision} = P(B|A) = P(A \cap B) / P(A)$$

예를 들어, 데이터베이스 전체에서 질의 영상과 같은 군에 속하는 영상의 개수(B)이 50이고, 특정 개수(A)

가 40이고, 데이터베이스 내에서 질의영상과 같은 군에 속하는 영상이면서 특정 개수 내에 속하는 영상(C)이 30개라고 하면 Recall은 30/50 혹은 0.60이 된다. 이것은 아래 그림 1에서 Ra/R가 된다. 그리고, Precision은 30/40 혹은 0.75가 된다. 즉, 그림 1의 C/A가 된다.

[3]에서는, 영상검색을 위한 특징벡터로

- (1) 인접 영역간의 상관 관계를 이용한 방법(SM1)과,
- (2) 영상 내 특정 영역들에 가중치를 부여하여 비교하는 방법(SM2), 그리고,
- (3) (2)에서의 가중치를 한 영역의 중심으로부터 거리에 따라 다르게 적용하는 방법(SM3)을 제안하고, 기존에 나와 있는 Color Histogram을 이용한 방법(CH), Correlogram을 이용하는 방법(Corr)과 Recall을 이용하여 결과를 비교하였다.

표 1. 특정 개수에 따른 Recall의 평균값

Scope	CH	Corr.	SM(1)	SM(2)	SM(3)
5	0.12	0.29	0.70	0.80	0.95
50	0.23	0.44	0.79	0.85	1.0
500	0.44	0.57	0.85	0.95	1.0

표 1은 [4]에서 언급한 결과로, 이 표를 근거로 자신들이 제안한 방법들(SM1, SM2, SM3)이 검색 효율이 높다고 주장한다.

또한 [5]에서는 특징벡터로 영상 내에 같은 색상들의 응집정도를 고려하는 방법[10](CCV : Color Coherence Vector)에 연속적인 세분화 과정을 거치는 방법(CCV/C)과 자기상관관계를 이용하는 방법(auto)을 제안하였다. 여기서도 자신들이 제안한 방법이 칼라 히스토그램(hist) 방법보다 효율적임을 주장하는 근거로써 아래 표 2와 같이 Recall 값을 비교했다.

표 2. 특정 개수에 따른 Recall 결과들

Scope	Recall		
	hist	CCV/C	auto
10	0.14	0.19	0.24
30	0.19	0.19	0.38
50	0.19	0.24	0.57

표 1과 표 2의 결과를 보면, 임의의 특정 개수에 대해, 거리의 크기와 상관없이 속하는 여부만을 비율로 나타낼 뿐, 해당 특정 개수 범위 전체에 걸친 거리의 분포를 전혀 고려하지 않기 때문에 방법을 비교하는데 사용할 수 있는 많은 정보를 이용하지 않고 있다. 즉, 표 1에서 표시된 수치는 단지 특정 개수의 범위에서 질의 영상과 관계 있는 영상이 몇 개나 검출되었는지를 알 수 있을 뿐이지 거리들의 전체적인 분포를 알 수는 없다. 영상들의 검색 결과 등급이, 특정 개수 범위 내에서 앞으로 물려있을 수도 있고, 고르게 퍼져있을 수도 있고, 모두 뒤쪽에 물려 있을 수도 있다. 예를 들어, 검출되어 올림차순으로 정렬된 특징벡터간의 거리 값들 중, 질의영상이 속해있는 영상 군에 속하는 영상의 개수가 5개라면, 표 1의 특정 개수=5에 대해 나온 결과가 의미하는 바는 각 방법에 대해 0.6개(CII), 1.45개(Corr), 3.5개(SM(1)), 4개(SM(2)), 4.75개(SM(3))의 영상이 특정 개수 내에 검색된 결과 중에 포함된다는 것이다. 여기서 5개가 모두 포함된 SM(3)의 경우와 하나도 포함되지 않은 CII의 경우를 제외하고, Corr인 경우에 포함된 1개가 첫 번째로 검색된 것이고, SM(1)과 SM(2)에서의 3개와 4개가 두 번째 이하로 검색되었다면 SM(1) 이나, SM(2)가 더 낫다고 말할 수 없다. 이것은 Recall/Precision 방법이 주어진 샘플인, 특징벡터들간의 거리 값의 순위들의 분포를 고려하지 않기 때문에 각 순위 값의 위치들을 고려하지 못하고 단지 일정 범위 내에 속하는 원소의 개수만을 고려하기 때문에 그렇게 얻어진 샘플들에 대한 비교를 가지고 어떤 방법이 더 좋다고 분명히 말할 수 없다고 본다. 따라서 샘플들의 분포를 고려하여 다른 방법들의 검색 유효성을 비교할 수 있는 방법으로 다음절에서 Wilcoxon rank Test를 설명하고, 내용기반 영상검색 결과들에 대한 유효성 검사에의 적용을 제안하고자 한다.

3. Wilcoxon rank Test

3.1 기본 개념

비교하고자 하는 두 개의, 특징벡터를 추출하는 방법들에 의해 구해진 데이터베이스 내의 각 영상들의 특징 벡터들과 질의 영상의 특징벡터간의 거리들을 각각 샘플 A와 샘플 B라고 하자. 여기서 A의 원소의 개수는 N개라 하고, B의 원소의 개수는 M개라고 하자. 두 샘플들을 (M+N)개의 원소로 이루어진 하나의 샘플로 합하여 가장 작은 값부터 가장 큰 값까지 등수를 부여한다. 이 때 각각의 원소에 부여된 등위를 그 값의 순위(rank)라고 한다. 이

기서, 두 표본에 해당하는 모든 원소들은 연속분포로 생성되었다고 가정하면 전체 순위의 합은

$$S = \sum_{i=1}^{M+N} R_i = 1+2+\dots+(M+N) = \frac{(M+N)(M+N+1)}{2}$$

가 되고 순위에 대한 평균과 분산은 다음과 같다.

$$E = \frac{S}{M+N} = \frac{M+N+1}{2}$$

$$V = \frac{1}{M+N} \sum_{i=1}^{M+N} (R_i - E)^2 = \frac{(M+N+1)(-M-N+1)}{12}$$

예를 들어, 표본 A={4, 6, 1, 7, 2}이고, 표본 B={10, 3, 5, 8, 9}이라면, 두 표본들이 정렬되어 병합된 표본 C는 다음과 같다.

$$C = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

이 표본의 평균은 5.5이고, 분산은 9.167이다. 이 경우는 각 표본의 값이 병합된 표본에서의 순위에 해당되므로, 표본 C내에서 표본 A 원소들의 순위 값의 평균은 4이고 분산은 6.5이며, 표본 B 원소들의 순위 값의 평균은 7이고 분산은 8.5가 된다. 표본 B의 평균 순위가 표본 A의 평균 순위보다 상대적으로 큰 것을 볼 수 있다. 이것은 A에 적용된 검색 방법이 B의 방법보다 더 나은 검색 결과를 갖는다고 말할 수 있다.

3.2 적용

3.2.1 샘플

표 3은 내용기반 영상검색의 특징벡터 추출 방법 A, B, C를 적용한 후, 질의영상의 특징벡터와의 거리 값을 계산하여 올림차순으로 정리하여, 각 순위에 해당하는 영상번호를 적었다. 여기서 3가지 방법 모두에 대해 사용된 질의영상(Q)은 영상번호가 1인 영상이다. 그리고, 질의영상이 속한 영상 군에 속하는 영상들은 영상 1부터 18까지에 해당되는 영상들이다. 표 3의 방법 A, B, C 모두에 대해, 첫 번째 열이 나타내는 것은, 각 순위에 해당되는 영상번호이고, 두 번째 열의 값은 그 영상의 특징벡터와 질의영상의 특징벡터와의 거리 값들이다. 세 번째 열의 값은, A, B, C 각 방법들에 각각 부여된 거리 값들을 통합하여 순위를 부여했을 때의 순위 값들이다. 2개의 순위 값이 같을 때는 다음으로 매겨질 순위 값 2개의 평균을 두 개에 동일하게 부여한다.

표 3. 영상검색 결과 비교 예.

Rank	특징벡터 추출방법								
	A			B			C		
1	1	0.0	1.5	1	0.7	8.5	1	0.0	1.5
2	2	0.1	3	2	0.8	11	4	0.3	4.5
3	18	0.3	4.5	18	0.9	13.5	5	0.6	7
4	17	0.5	6	12	1.5	19	3	0.7	8.5
5	12	0.8	11	15	1.7	23	7	0.8	11
6	13	1.1	16	3	2.0	26.5	2	0.9	13.5
7	14	1.2	17	11	2.2	28	6	1.0	15
8	15	1.5	19	13	2.4	30	8	1.5	19
9	11	1.7	23	17	2.7	34	18	1.6	21
10	3	1.9	25	10	3.4	42	11	1.7	23
11	10	2.3	29	14	3.7	44.5	9	1.8	24
12	5	2.5	31	4	3.9	46.5	17	2.0	26.5
13	9	2.6	32	9	4.4	52	64	2.7	34
14	4	2.7	34	5	4.5	53	65	2.9	37.5
15	6	2.8	36	6	4.7	55.5	66	3.0	39
16	7	2.9	37.5	7	5.5	64	67	3.2	40
17	8	3.3	41	8	5.6	65.5	10	4.0	48
18	16	3.5	43	19	5.7	67	14	4.1	49
19	33	3.7	44.5	22	5.8	69	15	4.2	50.5
20	34	3.9	46.5	24	6.4	75	12	4.7	55.5
21	35	4.2	50.5	23	6.5	76	13	4.8	57
22	37	4.6	54	25	6.6	77	16	4.9	58.5
23	42	4.9	58.5	20	6.7	78.5	25	5.0	60
24	39	5.1	61.5	26	7.5	85.5	39	5.1	61.5
25	38	5.3	63	27	7.6	87	23	5.8	69
26	36	5.6	65.5	21	7.7	88	24	5.9	71
27	40	5.8	69	89	8.5	90	30	6.0	72.5
28	41	6.0	72.5	16	8.6	91	29	7.0	80
29	32	6.3	74	64	8.7	92	33	7.2	81
30	20	6.7	78.5	65	9.2	93	34	7.3	82
31	27	7.4	83.5	66	9.4	94	37	7.4	83.5
32	21	8.0	89	67	9.9	95	39	7.5	85.5

3.2.2 Wilcoxon 순위방법의 적용

표 4는 표 3에서의 샘플에 대한 Recall을 계산한 것이다.

표 4. 표 3에 대한 Recall 결과들

Scope	Recall		
	A	B	C
5	5/18	5/18	5/18
10	10/18	10/18	10/18
15	15/18	15/18	12/18
20	18/18	18/18	16/18
25	18/18	18/18	18/18
30	18/18	18/18	18/18

A와 B에 대한 Recall의 결과는 동일하고, C에 대한 결과가 가장 안 좋게 나타난다. 표 5는 표 3에서 통합된 순위로부터 A, B, C 각각에 할당된 순위 값들에 대한 평균과 분산을 계산한 것이다. 즉, Wilcoxon 순위방법을 이용한 것이다.([2]를 참조)

표 5. Table 4의 A, B, C에 대한 통합 rank들 중 각각에 해당되는 값들에 대한 평균과 분산

	A	B	C
평균	.41.250	58.578	43.422
분산	633.355	793.857	711.115

표5에서의 결과를 통해, A, B, C에 결과적으로 할당된 순위들을 비교하면 아래와 같이 나타낼 수 있다.

$$\text{rank}(A) < \text{rank}(C) < \text{rank}(B)$$

즉, A방법에 의한 A에 속하는 순위들이 상대적으로 B나 C에 할당된 순위들보다 작고, 그 다음으로 C에 속하는 순위들 그리고 B에 속하는 순위들이 상대적으로 큰 값들을 알 수 있다. 즉, A를 사용했을 때 질의영상이 속한 영상군의 영상들을 가장 잘 검색하였고 그 다음으로 C 그리고 마지막으로 B 방법에 의한 검색 결과가 검색한 비슷한 영상의 개수가 가장 적음을 나타낸다. 이 결과를 표 4의 Recall에 의한 결과와 비교해보면, A와 B에 의한 결과가 동일하며, C가 상대적으로 좋지 않다고 결과를 보인 Recall에 의한 결과를 신뢰하는데 문제가 있고, 제안한 방법으로 영상검색 방법들에 대한 비교 및 평가를 더 정확하게 수행할 수 있음을 알 수 있다.

5. 결론

순위를 이용하여 내용 기반 영상 검색의 결과들에 대한 효율성을 평가, 비교하는 방법을 제안하였다. 다른 연구들에서 주로 사용하는 Recall/Precision 방법은 비교 대상이 되는 두 샘플간에 전체적인 경향을 제공하기보다 주어진 범위 내에 속한 영상들의 개수만으로 비교를 수행한다. 따라서 정확한 평가를 제공하지 못한다. 제안한 방법을 이용하여 이러한 문제를 해결할 수 있다고 본다.

[참고문헌]

- [1] Adjeroh, D. A., Lee, M.C., "An Occupancy Model for Image Retrieval and Similarity Evaluation," *IEEE Trans. on Image Processing*, VOL. 9, NO. 1, pp. 120-131, Jan. 2000.
- [2] Conover, W. J., *Practical Nonparametric Statistics*, Third Edition, John Wiley & Sons, Inc., 1999.
- [3] Chahir, Y., "Searching Images on the Basis of Color Homogeneous Objects and their Spatial Relationship," *Journal of Visual Communication and Image Representation 11*, pp. 302-326, 2000.
- [4] Hua, K. A., Vu, K. and Oh, J. H., "SamMatch: A Flexible and Efficient Sampling-Based Image Retrieval Techniques for Large Image Database," *ACM Multimedia*, (1999), 225-234.
- [5] Huang, J., Kumar .S. R., Mitra, M., Zhu, W., and Zabih, R. "Image Indexing Using Color Correlogram," *In Proceedings of Computer Vision and Pattern Recognition*. pp. 762-768, 1997.
- [6] Heuer, J., Sanahuja, F., Kaup, A., "Visual Feature Discrimination versus Compression Ratio for Polygonal Shape Descriptors," *Proceeding of SPIE Vol. 4210*, pp. 82-93, 2000.
- [7] Kim, Y. S., Kim, W. Y., "Content-based trademark retrieval system using a visually salient feature," *Image and Vision Computing 16*, pp. 931-939, 1998.
- [8] Lecce, V. D., Guerriero, A., "An Evaluation of the Effectiveness of Image Features for Image Retrieval," *Journal of Visual Communication and Image Representation 10*, pp.351-362, 1999.
- [9] Mehtre, B. M., Narasimhalu, K. A. D., Man, G. C., "Color matching for image retrieval," *Pattern Recog. Lett. 16*, pp. 325-331, 1995.
- [10] Pass, G., Zabih, R., Miller, J., "Comparing Images Using Color Coherence Vectors," *ACM Multimedia*, pp. 65-73, 1996.
- [11] Smith, J. R., *Integrated Spatial and Feature Image System : Retrieval, Analysis and Compression*, Ph. D Thesis, Columbia Univ. 1998